ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Linear Algebra and its Applications 390 (2004) 197–208
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Tridiagonal doubly stochastic matrices

Geir Dahl*

Centre of Mathematics for Applications, University of Oslo, c/o Dept. of Mathematics, P.O. Box 1053, Blindern, NO-0316 Oslo, Norway

Received 29 January 2004; accepted 20 April 2004

Submitted by R.A. Brualdi

Abstract

We study the facial structure of the polytope Ω_n^{t} in $\mathbb{R}^{n\times n}$ consisting of the tridiagonal doubly stochastic matrices of order n. We also discuss some subclasses of Ω_n^{t} with focus on spectral properties and rank formulas. Finally we discuss a connection to majorization.

Keywords: Doubly stochastic matrix; Birkhoff polytope; Eigenvalue; Random walk; Majorization

1. Introduction

A (real) $n \times n$ matrix A is *doubly stochastic* if it is nonnegative and all its row and column sums are one. The *Birkhoff polytope*, denoted by Ω_n, consists of all doubly stochastic matrices of order n. A well-known theorem of Birkhoff and von Neumann (see [3]) states that Ω_n is the convex hull of all permutation matrices of order n. In this paper we discuss the subclass of Ω_n consisting of the tridiagonal doubly stochastic matrices and the corresponding subpolytope

$$\Omega_n^{\mathrm{t}} = \{A \in \Omega_n : A \text{ is tridiagonal}\}$$

of the Birkhoff polytope. We call Ω_n^{t} the *tridiagonal Birkhoff polytope*. Ω_n^{t} is a face of Ω_n and the structure of this face is investigated in the next section. Throughout the paper we assume that $n \geqslant 2$.

The permanent of tridiagonal doubly stochastic matrices was investigated in [7] and it was shown that the minimum permanent in this class is $1/2^{n-1}$ (where n

* Tel.: +47-22-85-5835; fax: +47-22-85-4349.
E-mail address: geird@math.uio.no (G. Dahl).

0024-3795/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2004.04.017

denotes the order of the matrices). We remark that this result may also be derived from a related result in [4].

Tridiagonal doubly stochastic matrices arise in connection with random walks on the integers $\{1, 2, \ldots, n\}$ where (i) in a single transition from an integer i the process (say, a person) either stays in i or moves to an adjacent integer, and (ii) the transition probabilities are symmetric in the sense that $p_{i,i+1} = p_{i+1,i}$ $(1 \leqslant i \leqslant n-1)$. We return to this example in Section 4.

The notation in this paper is as follows. An all zeros matrix is denoted by O, and we let J_n (or simply J) denote the all ones square matrix of order n. For a matrix (or vector) A we write $A \geqslant O$ if A is (componentwise) nonnegative. As usual the components of a vector $x \in \mathbb{R}^n$ are denoted by x_i, so $x = (x_1, x_2, \ldots, x_n)$. The cardinality of a finite set S is denoted by $|S|$.

2. The polytope Ω_n^{t}

We first describe a representation of all matrices in Ω_n^{t}. Define the polytope

$$P_n = \{\mu \in \mathbb{R}^{n-1} : \mu \geqslant O, \mu_i + \mu_{i+1} \leqslant 1 \ (1 \leqslant i \leqslant n-2)\} \tag{1}$$

in $\mathbb{R}^{n-1}$ for $n \geqslant 3$. We also define $P_2 = [0, 1]$. For each vector $\mu \in \mathbb{R}^{n-1}$ we define the associated $n \times n$ matrix

$$A_\mu = \begin{bmatrix} 1-\mu_1 & \mu_1 & 0 & 0 & \cdots & 0 \\ \mu_1 & 1-\mu_1-\mu_2 & \mu_2 & 0 & \cdots & 0 \\ 0 & \mu_2 & 1-\mu_2-\mu_3 & \mu_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mu_{n-2} & 1-\mu_{n-2}-\mu_{n-1} & \mu_{n-1} \\ 0 & 0 & \cdots & \cdots & \mu_{n-1} & 1-\mu_{n-1} \end{bmatrix}.$$

So this is a symmetric matrix and its subdiagonal is equal to μ. If $\mu \in P_n$, then the matrix A_μ is doubly stochastic and tridiagonal, i.e., $A_\mu \in \Omega_n^{\mathrm{t}}$. A useful fact is that every matrix in Ω_n has the form A_μ for some $\mu \in P_n$.

Proposition 1

$$\Omega_n^{\mathrm{t}} = \{A_\mu : \mu \in P_n\}.$$

Proof. The inclusion $\{A_\mu : \mu \in P_n\} \subseteq \Omega_n^{\mathrm{t}}$ is clear. For the opposite inclusion, consider a tridiagonal doubly stochastic matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & 0 & \cdots & 0 \\ 0 & a_{32} & a_{33} & a_{34} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & a_{n\,n-1} & a_{nn} \end{bmatrix}.$$

Define $\mu_i = a_{i\,i+1}$ for $i = 1, 2, \ldots, n-1$ and let $\mu = (\mu_1, \mu_2, \ldots, \mu_{n-1})$. We now verify that $A = A_\mu$. As A is doubly stochastic, $a_{11} = 1 - \mu_1$ and $a_{21} = \mu_1$ as desired. Assume, for a given i, that $a_{i\,i-1} = \mu_{i-1}$. Since the ith row sum is one and $a_{i\,i+1} = \mu_i$, we obtain $a_{ii} = 1 - \mu_{i-1} - \mu_i$. Similarly, by considering the ith column, we calculate $a_{i+1\,i} = 1 - a_{ii} - a_{i-1\,i} = 1 - (1 - \mu_{i-1} - \mu_i) - \mu_{i-1} = \mu_i$. It follows, by induction, that $A = A_\mu$. $\square$

Thus, every matrix in Ω_n^{t} is determined by its superdiagonal (or subdiagonal). Moreover we see that P_n and Ω_n^{t} are affinely isomorphic. This means that the polyhedral structure of the tridiagonal Birkhoff polytope is found directly from the corresponding structure of P_n.

Let f_n denote the nth Fibonacci number. So $f_1 = f_2 = 1$ and $f_n = f_{n-1} + f_{n-2}$ for each $n \geqslant 3$. We recall that f_n is given explicitly as

$$f_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$$

(see e.g. [2]). Polyhedral properties of the tridiagonal Birkhoff polytope are collected in the following theorem where we use the notation $K = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $J = [1]$.

Theorem 2. (i) *Ω_n^{t} is a polytope in $\mathbb{R}^{n\times n}$ of dimension $n-1$ with f_{n+1} vertices.*

(ii) *Its vertex set consists of all tridiagonal permutation matrices; these are the matrices of order n that can be written as a direct sum*

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_t \tag{2}$$

where each matrix A_i $(i \leqslant t)$, hereafter called a block, *equals either J or K.*

(iii) *Consider a vertex A as in* (2). *Then each adjacent vertex of A is obtained from A by either* (a) *interchanging a sequence of consecutive blocks $J, K, K, \ldots, K$ (with $t \geqslant 1$ Ks) and the sequence $K, K, \ldots, K, J$ (with t Ks), or* (b) *by interchanging a sequence of consecutive blocks $K, K, \ldots, K$ (with $t \geqslant 1$ Ks) and the sequence $J, K, K, \ldots, K, J$ (with $t-1$ Ks).*

Proof. Since Ω_n^{t} and P_n are affinely isomorphic, we may prove the theorem by considering P_n. Clearly, P_n has dimension $n-1$, since it contains all coordinate vectors and the zero vector. Therefore, Ω_n^{t} has dimension $n-1$. Using the extreme point property it is easy to verify that P_n has only integral vertices, i.e., all components are integers. It follows that the vertex set of P_n, denoted by V_n, consists of all (0, 1)-vectors μ of length $n-1$ not having two consecutive 1s. (Actually, P_n is the stable set polytope associated with the graph which is a path of length $n-1$.) The corresponding matrices A_μ are the direct sum of matrices in the set $\{J, K\}$. We next determine the cardinality of the vertex set V_n. There is a bijection between $\{\mu \in V_n : \mu_{n-1} = 0\}$ and V_{n-1}; it is obtained by dropping the last component of $\mu \in V_n$ (as $\mu_{n-1} = 0$). Similarly, there is a bijection between $\{\mu \in V_n : \mu_{n-1} = 1\}$ and V_{n-2};

it is obtained by dropping the last two components of $\mu \in V_n$ (as $\mu_{n-1} = 1$ and $\mu_{n-2} = 0$). It follows that $|V_n| = |V_{n-1}| + |V_{n-2}|$ for $n \geqslant 4$. Clearly, $|V_2| = 2$ and $|V_3| = 3$. This means that the cardinalities $|V_n|$ ($n \geqslant 2$) are given by the Fibonacci numbers: $|V_n| = f_{n+1}$ for each n. This proves (i) and (ii).

To prove (iii) consider two distinct vertices μ, μ' of P_n, and let $S = \{j : \mu_j = 1\}$, $S' = \{j : \mu'_j = 1\}$. We may write

$$S \,\Delta\, S' = I_1 \cup I_2 \cup \cdots \cup I_p,$$

where $I_r = \{i_r, i_r + 1, \ldots, j_r\}$ for some integers $i_r \leqslant j_r$ ($r \leqslant p$) with $i_{r+1} \geqslant j_r + 2$ ($r \leqslant p - 1$).

Claim. *μ and μ' are adjacent if and only if $p = 1$, i.e., $S \,\Delta\, S'$ is an (integer) interval.*

Assume first that $p \geqslant 2$. Let $\gamma \in \mathbb{R}^{n-1}$ be the vector obtained from μ by letting $\gamma_j = 1 - \mu_j$ for each $j \in I_1$. Similarly, let $\gamma' \in \mathbb{R}^{n-1}$ be obtained from μ' by letting $\gamma'_j = 1 - \mu'_j$ for each $j \in I_1$. Then $\mu, \mu', \gamma, \gamma'$ are four distinct vertices of P_n satisfying $(1/2)(\mu + \mu') = (1/2)(\gamma + \gamma')$ which implies that the smallest face of P_n containing μ and μ' has dimension at least two. Thus, if $p \geqslant 2$, then μ and μ' are not adjacent. Next, assume that $p = 1$ and define the vector $w \in \mathbb{R}^{n-1}$ as follows: $w_j = n^2$ when $j \in S \cap S'$, $w_j = -1$ when $j \notin S \cup S'$, $w_j = |S \setminus S'|$ when $j \in S' \setminus S$ and, finally, $w_j = |S' \setminus S|$ when $j \in S \setminus S'$. Then one can check that the only vertices of P_n that maximize the linear function $w^{\mathrm{T}} z$ for $z \in P_n$ are μ and μ'. This implies that these two vertices are adjacent on P_n. This proves our claim, and (iii) follows by translating this adjacency characterization into matrix language. □

Let $G(\Omega_n^{\mathrm{t}})$ denote the *graph of* Ω_n^{t} (or 1-skeleton), i.e., the vertices and edges of the graph $G(\Omega_n^{\mathrm{t}})$ correspond to the vertices and edges of the polytope Ω_n^{t}. In Theorem 2 the vertices and edges of Ω_n^{t} were described. We now determine the diameter of $G(\Omega_n^{\mathrm{t}})$ which is defined as the maximum of $d(u, v)$ taken over all pairs u, v of vertices, where $d(u, v)$ is the smallest number of edges in a path between u and v in $G(\Omega_n^{\mathrm{t}})$.

Theorem 3. *The diameter of $G(\Omega_n^{\mathrm{t}})$ equals $\lfloor n/2 \rfloor$.*

Proof. Consider two distinct vertices μ, μ' of P_n. As in the proof of Theorem 2 we let $S = \{j : \mu_j = 1\}$, $S' = \{j : \mu'_j = 1\}$ so

$$S \,\Delta\, S' = I_1 \cup I_2 \cup \cdots \cup I_p.$$

Since each I_t is nonempty and consecutive intervals are nonadjacent, it follows that $p + (p - 1) \leqslant n - 1$. So $p \leqslant \lfloor n/2 \rfloor$. We may now find a path

$$Q : \mu = \mu^{(0)}, \mu^{(1)}, \ldots, \mu^{(p)} = \mu'$$

of length p in $G(\Omega_n^{\mathrm{t}})$ where $\mu^{(t)}$ is obtained from $\mu^{(t-1)}$ by complementing zeros and ones for indices in I_t $(t \leqslant p)$. We see from the adjacency characterization of Theorem 2 that $\mu^{(t-1)}$ and $\mu^{(t)}$ are adjacent. Thus, $G(\Omega_n^{\mathrm{t}})$ contains a path between any pair of vertices of length $p \leqslant \lfloor n/2 \rfloor$, and therefore the diameter of $G(\Omega_n^{\mathrm{t}})$ is at most $\lfloor n/2 \rfloor$. To prove equality here consider first the case when n is even, say $n = 2k$. The distance (in $G(\Omega_n^{\mathrm{t}})$) between the matrices $A = J \oplus J \oplus \cdots \oplus J$ (with $2k$ Js) and $B = K \oplus K \oplus \cdots \oplus K$ (with k Ks) is at least k since for any two adjacent vertices their number of Ks differ by at most one (see Theorem 2). If n is odd, $n = 2k + 1$, we consider the matrices obtained from A and B above by adding a J block (at the end) and conclude that their distance is at least $k = \lfloor n/2 \rfloor$ as desired. $\square$

We conclude this section by some observations concerning optimization over the set Ω_n^{t}. Let C be a given square matrix of order n. The well-known *assignment problem* is to maximize a linear function $\langle C, A\rangle = \sum_{i,j} c_{ij}a_{ij}$ over all permutation matrices A. Equivalently, we may here maximize over the set Ω_n of doubly stochastic matrices; this follows from Birkhoff's theorem as the objective function is linear. Consider now the more restricted problem of maximizing $\langle C, A\rangle$ over the tridiagonal permutation matrices A, or equivalently, over $A \in \Omega_n^{\mathrm{t}}$. We may then assume that C is also tridiagonal. By using the relation between Ω_n^{t} and the polytope P_n (see Proposition 1) our problem reduces to a linear optimization problem over P_n (where the d_js are calculated from C):

$$\max \left\{ \sum_{j=1}^{n-1} d_j\mu_j : \mu \in P_n \right\}. \tag{3}$$

Now, this problem may be solved by dynamic programming as follows. Define $v_k = \max\{\sum_{j=1}^{k} d_j\mu_j : \mu_j + \mu_{j+1} \leqslant 1\ (j \leqslant k-1), \mu_1, \ldots, \mu_k \geqslant 0\}$ and note that v_{n-1} is the optimal value of (3). The algorithm is: (i) $v_1 = \max\{0, d_1\}$, $v_2 = \max\{v_1, d_2\}$, (ii) for $k = 3, 4, \ldots, n-1$ let $v_k = \max\{v_{k-1}, v_{k-2} + d_k\}$. This simple algorithm is linear, and by storing some more information we also find an optimal solution $\mu_1, \mu_2, \ldots, \mu_{n-1}$.

3. Diagonally dominant matrices in Ω_n^{t}

In this section we consider the tridiagonal doubly stochastic matrices that are diagonally dominant. Recall that a matrix $A = [a_{ij}]$ of order n is called (*row*) *diagonally dominant* if $|a_{ii}| \geqslant \sum_{j:j\neq i} |a_{ij}|$. If all these inequalities are strict, then A is called *strictly* (*row*) *diagonally dominant*, and it is well-known that this property implies that A is nonsingular.

Let

$$\Omega_n^{\mathrm{t,d}} = \{A \in \Omega_n^{\mathrm{t}} : A \text{ is diagonally dominant}\}$$

and note that, since each $A \in \Omega_n^{\mathrm{t}}$ is symmetric, we need not distinguish between row and column diagonally dominance. We remark that every matrix A in $\Omega_n^{\mathrm{t,d}}$ is also *completely positive*, i.e., $A = BB^{\mathrm{T}}$ for some nonnegative $n \times k$ matrix B. Moreover, the smallest k in such a representation (called the cp-rank of A) is equal to the rank of A. We refer to the recent book [1] for a survey of completely positive matrices. These two facts concerning matrices in Ω_n^{t} follow from the general theory in [1], or a direct verification is also possible.

The following theorem shows that $\Omega_n^{\mathrm{t,d}}$ is very similar to Ω_n^{t}. In the following discussion we define $\mu_0 = \mu_n = 0$.

Theorem 4
(i) $\Omega_n^{\mathrm{t,d}}$ *is a subpolytope of* Ω_n^{t}.
(ii) $\Omega_n^{\mathrm{t,d}} = \{A_\mu : \mu \geqslant O, \mu_i + \mu_{i+1} \leqslant 1/2 \ (i \leqslant n-2)\} = \{A_\mu : \mu \in (1/2)P_n\}$.
(iii) *The vertex set of* $\Omega_n^{\mathrm{t,d}}$ *consists of the matrices of order n that may be written as a direct sum of matrices in the set* $\{J_1, (1/2)J_2\}$.

Proof. The matrix A_μ is diagonally dominant if and only if $1 - (\mu_{i-1} + \mu_i) \geqslant \mu_{i-1} + \mu_i \ (1 \leqslant i \leqslant n)$, i.e., iff $\mu_{i-1} + \mu_i \leqslant 1/2 \ (1 \leqslant i \leqslant n)$. This implies (ii) and also (i). To see (iii) we recall from the proof of Theorem 2 that the vertex set of P_n consists of all (0, 1)-vectors μ (of length $n-1$) not having two consecutive 1s. So the vertices of the polytope $(1/2)P_n$ are the (0, 1/2)-vectors not having two consecutive $\frac{1}{2}$s. This implies (iii). □

We now investigate the rank of the matrices in the class $\Omega_n^{\mathrm{t,d}}$.

Theorem 5. *Let* $A_\mu \in \Omega_n^{\mathrm{t,d}}$. *Then*

$$\mathrm{rank}(A_\mu) = n - |\{i : \mu_i = 1/2\}|.$$

In particular, $\mathrm{rank}(A_\mu) \geqslant \lfloor n/2 \rfloor$.

Proof. Consider a matrix $A_\mu \in \Omega_n^{\mathrm{t,d}}$, so $\mu \in (1/2)P_n$. If $\mu_i = 0$, for some i with $1 \leqslant i \leqslant n-1$, then A_μ is the direct sum of two matrices of order i and $n-i$, respectively. Therefore, since the rank of a direct sum of some matrices is the sum of the ranks of these matrices, it suffices to prove the result for the case when $\mu_i > 0$ $(1 \leqslant i \leqslant n-1)$. There are two possibilities. First, if $\mu_i = 1/2$ for some i, then it follows from the diagonal dominance that $\mu_{i-1} = \mu_{i+1} = 0$. This implies that $n = 2$ and that $A_\mu = (1/2)J_2$ and the rank formula holds. Alternatively, when $\mu_i < 1/2$ for each i, then $a_{11} = 1 - \mu_1 > \mu_1 = \sum_{j=2}^{n} a_{1j}$ and this combined with the diagonal dominance of A_μ (and that each $\mu_i > 0$) implies that A_μ is nonsingular (confer Theorem 3.6.8 in [3]). This implies the rank formula. The lower bound on the rank is due to the fact μ does not contain two consecutive components that are 1/2 whenever $\mu \in (1/2)P_n$. □

Thus, we have a simple formula for the rank of matrices in the subclass $\Omega^{\mathrm{t,d}}$. On the other hand, it is not as straightforward to determine the rank of a matrix $A \in \Omega_n^{\mathrm{t}} \setminus \Omega_n^{\mathrm{t,d}}$. A is then a direct sum of matrices A_i, say of order k_i, for which the corresponding μ_is are positive. Clearly each A_i has rank k_i or $k_1 - 1$, and to decide which is the case one can solve a triangular linear system (in order to determine if the first column of A_i lies in the span of the other columns). The nonsingularity of each A_i may be expressed by a polynomial equation in the μ_js, but it seems very complicated.

4. Matrices in $\Omega^{\mathrm{t,d}}$ with constant subdiagonal

Consider the subpolytope

$$\Omega_n^{t,=} = \{A_\mu \in \Omega_n^{\mathrm{t}} : \mu_1 = \mu_2 = \cdots = \mu_{n-1}\}$$

of Ω_n^{t}. The corresponding subpolytope of P_n (in the space of the μ-variables) is simply the line segment $[O, (1/2)e]$. Note that a matrix in $\Omega_n^{t,=}$ may or may not be diagonally dominant.

Our main goal is to find explicitly all eigenvalues and corresponding eigenvectors for every matrix $A_\mu \in \Omega_n^{t,=}$. This is done by solving certain difference equations. A similar approach for finding eigenvalues and eigenvectors of tridiagonal Toeplitz matrices may be found in e.g. [6,10] (the latter reference also treats an extension to so-called pseudo-Toeplitz matrices).

Let $0 \leqslant x \leqslant 1/2$ and consider the (general) matrix

$$A_x = \begin{bmatrix} 1-x & x & 0 & 0 & \cdots & 0 \\ x & 1-2x & x & 0 & \cdots & 0 \\ 0 & x & 1-2x & x & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & x & 1-2x & x \\ 0 & 0 & \cdots & \cdots & x & 1-x \end{bmatrix}$$

in $\Omega_n^{t,=}$. Observe that $A_x = I - x \cdot W_n$ where W_n is the $n \times n$ matrix

$$W_n = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & \cdots & \cdots & -1 & 1 \end{bmatrix}.$$

It follows that the eigenvalues of A_x are $1 - x\lambda$ where λ is an eigenvalue of W_n. The corresponding eigenvectors are the same. Thus, we need to determine the spectrum of W_n. Note that W_n resembles the tridiagonal Toeplitz matrix

$$T_n = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & \cdots & \cdots & -1 & 2 \end{bmatrix},$$

which has eigenvalues $2-2\cos\left(\frac{j\pi}{n+1}\right)$ and corresponding eigenvector $s_j \in \mathbb{R}^n$ given by

$$s_j = \left(\sin\left(\frac{j\pi}{n+1}\right), \sin\left(\frac{2j\pi}{n+1}\right), \ldots, \sin\left(\frac{nj\pi}{n+1}\right)\right) \quad \text{for } 1 \leqslant j \leqslant n$$

(see e.g. [10]). We now show that the eigenvalues of W_n are the eigenvalues of T_{n-1} plus the eigenvalue 0 (so W_n is singular).

Theorem 6. *The eigenvalues of W_n are*

$$2 - 2\cos(j\pi/n) \quad (0 \leqslant j \leqslant n-1).$$

In particular W_n is singular. The corresponding (orthogonal) eigenvectors are

$$(2\cos(\pi j(k-1/2)/n))_{k=1}^n \quad (0 \leqslant j \leqslant n-1).$$

Proof. Let λ be an eigenvalue and y a corresponding eigenvector of W_n. The eigenvector equation $(W_n - \lambda I)y = O$ may then be written as

$$-y_{k-1} + (2-\lambda)y_k - y_{k+1} = 0 \quad (1 \leqslant k \leqslant n) \tag{4}$$

where $y_0 := y_1$ and $y_{n+1} := y_n$. This is a linear second order difference equation with rather special boundary conditions. The corresponding characteristic equation $z^2 + (\lambda-2)z + 1$ has solutions $r_1, r_2 = (1/2)(2-\lambda) \pm \sqrt{(\lambda-2)^2 - 4}$. Consider first the case when the roots coincide, i.e., when λ is 0 or 4. If $\lambda = 4$, then $r_1 = r_2 = -1$ and the general solution of (4) is $y_k = (\alpha + \beta k)(-1)^k$ where α, β are constants. It is easy to see that the boundary conditions lead to a contradictions in this case (we get from $y_0 = y_1$ that $\beta = 2\alpha$, and then the second boundary condition $y_n = y_{n+1}$ has no solution). Therefore $\lambda = 4$ is not an eigenvalue of W_n. On the other hand, if $\lambda = 0$, then $r_1 = r_2 = 1$ and the solution of (4) is $y_k = \alpha + \beta k$. But $y_0 = y_1$ implies $\beta = 0$ so $y_k = \alpha$ for some constant α. This proves that 0 is an eigenvalue of W_n with corresponding eigenvector $(1, 1, \ldots, 1)$.

Consider next when the roots r_1 and r_2 are distinct. Since $z^2 + (\lambda-2)z + 1 = (z-r_1)(z-r_2)$ we must have $r_1 r_2 = 1$, i.e., $r_2 = r_1^{-1}$. Thus, the general solution of (4) is

$$y_k = \alpha r_1^k + \beta r_1^{-k}.$$

The condition $y_0 = y_1$ gives $\alpha + \beta = \alpha r_1 + \beta r_1^{-1}$. We may assume $r_1 \neq 1$ (for otherwise $\lambda = 0$; a case already discussed). Therefore $\beta = \alpha r_1$ so

$$y_k = \alpha\left(r_1^k + r_1^{1-k}\right).$$

Note that $\alpha \neq 0$; otherwise $y = O$ contradiction that y is an eigenvector. The boundary condition $y_n = y_{n+1}$ gives $r_1^n + r_1^{1-n} = r_1^{n+1} + r_1^{-n}$. Multiplying this equation by r_1^n and reorganizing terms gives $r_1^{2n}(1 - r_1) = 1 - r_1$. Therefore, as $r_1 \neq 1$, we must have $r_1^{2n} = 1$. So $r_1^2 = \mathrm{e}^{2\pi \mathrm{i} j/n}$ (where $\mathrm{i} = \sqrt{-1}$) for some j with $1 \leqslant j \leqslant n-1$ ($j = n$ is excluded as $r_1 \neq 1$). This shows that $r_1 = \mathrm{e}^{\pi \mathrm{i} j/n}$ and $r_2 = \mathrm{e}^{-\pi \mathrm{i} j/n}$. Moreover, using that $r_1 + r_2 = 2 - \lambda$ we obtain

$$\lambda = 2 - 2\cos(j\pi/n).$$

We have therefore found all the eigenvalues of W_n. An eigenvector corresponding to $\lambda = 2 - 2\cos(j\pi/n)$ (for fixed j) is $y = (y_k)$ given by

$$y_k = \alpha(\mathrm{e}^{\pi \mathrm{i} jk/n} + \mathrm{e}^{\pi \mathrm{i} j(1-k)/n}).$$

Letting $\alpha = \mathrm{e}^{-(1/2)\pi \mathrm{i} j/n}$ we get

$$y_k = \mathrm{e}^{\pi \mathrm{i} j(k-1/2)/n} + \mathrm{e}^{-\pi \mathrm{i} j(k-1/2)/n} = 2\cos(\pi j(k - 1/2)/n),$$

which gives the desired eigenvector. □

We may now determine the spectrum of A_x (where again $0 \leqslant x \leqslant 1/2$).

Corollary 7. *The eigenvalues of A_x are*

$$1 - 2x(1 - \cos(j\pi/n)) \quad (0 \leqslant j \leqslant n-1).$$

and the corresponding eigenvectors are described in Theorem 6.

Proof. This follows directly from Theorem 6 using the relation $A_x = I - x \cdot S$. □

The rank of A_x is determined in the next corollary.

Corollary 8. *If $x \in \{1/(2 - 2\cos(j\pi/n)) : \lceil n/3 \rceil \leqslant j \leqslant n-1\}$, then A_x has rank $n-1$. Otherwise A_x is nonsingular.*

Proof. The last $n-1$ columns of A_x are linearly independent, so A_x has rank $n-1$ or n. The result now follows from Corollary 7. □

Also note that the kernel of A_x (when A_x is singular) is known explicitly since we have determined a complete set of eigenvectors of A_x. The matrix $A_x \in \Omega_n^{t,=}$ is diagonally dominant if and only if $0 \leqslant x \leqslant 1/4$. From Corollary 7 it follows that A_x is positive semidefinite if and only if $0 \leqslant x \leqslant 1/(2 + 2\cos(\pi/n))$. Thus, when n is large, the class of positive semidefinite matrices in $\Omega_n^{t,=}$ is just "slightly larger" than the class of diagonally dominant matrices in $\Omega_n^{t,=}$.

For a general doubly stochastic matrix A the bound

$$|1 - \lambda| \geqslant 2(1 - \cos(\pi/n))\mu(A) \tag{5}$$

for eigenvalues $\lambda \neq 1$ of A was found by Fiedler. Here $\mu(A)$ is a measure of the irreducibility of A given by $\mu(A) = \min_M \sum_{i \in M} \sum_{j \notin M} a_{ij}$ where the minimum is taken over all nonempty strict subsets M of $\{1, 2, \ldots, n\}$. See [8] for a discussion of such estimates. It is interesting to check the quality of the bound (5) for matrices $A_x \in \Omega_n^{t,=}$, as we know the eigenvalues for these matrices. Let $A_x \in \Omega_n^{t,=}$. Then we find that $\mu(A_x) = x$. So if λ denotes the second largest eigenvalue of A_x, we get from Corollary 7 that $1 - \lambda = 2x(1 - \cos(\pi/n)) = 2(1 - \cos(\pi/n))\mu(A)$. This means that Fiedler's estimate is tight for this subclass $\Omega_n^{t,=}$ of the doubly stochastic matrices.

An application. We briefly discuss an application of Corollary 7 to Markov chains. Recall the specific random walk discussed in Section 1 and assume that the one-step transition matrix of the chain is A_x for some $x \in [0, 1/2]$. Thus, if p_{ij} is the probability of moving in one step from state i to state j, then we have $p_{ii+1} = p_{i+1i} = x$ $(1 \leqslant i \leqslant n-1)$, $p_{ii} = 1 - 2x$ $(2 \leqslant i \leqslant n-1)$, and $p_{11} = p_{nn} = 1 - x$ while all other p_{ij}s are zero. The explicit knowledge of the eigenvalues and eigenvectors of A_x, presented in Corollary 7, is very useful for analyzing the behavior of this random walk. To be specific, let U be the $n \times n$ matrix with the eigenvectors of A_x as its columns, and let D be the diagonal matrix with the associated eigenvalues along the diagonal. So $U^{\mathrm{T}} A_x U = D$ and since U is orthogonal we get $A_x^k = U D^k U^{\mathrm{T}}$ for each positive integer k. The (i, j)th entry of A_x^k equals the probability that the process goes from state i to state j in k transitions (see e.g. [5] for the theory of Markov chains). This means that one can calculate the k step transition probabilities (the powers of A_x) efficiently. Moreover, one can get explicit information about how fast the chain converges towards its stationary distribution (which is the uniform distribution as A_x is doubly stochastic) since we know all the eigenvalues.

5. $\Omega_n^{\mathbf{t}}$ and majorization

Doubly stochastic matrices are important in the area of majorization. For two vectors $x, y \in \mathbb{R}^n$ we say that x is majorized by y if $\sum_{i=1}^k x_{[i]} \leqslant \sum_{i=1}^k y_{[i]}$ for $k \leqslant n$ and where equality holds when $k = n$. Here $x_{[i]}$ denotes the ith largest component of x. A basic result here is a theorem of Hardy, Littlewood and Pólya saying that x is majorized by y if and only if there is a doubly stochastic matrix A such that $x = Ay$. For a discussion of this result and a strengthened result concerning restricted doubly stochastic matrices, so-called T-transforms, see [9].

Motivated by the mentioned theorem we now define a majorization concept which is stronger than ordinary majorization. Let $x, y \in \mathbb{R}^n$ be *monotone* vectors, i.e., the components are nonincreasing. We say that x is *tridiagonally majorized* by y if there is a tridiagonal doubly stochastic matrix A such that $x = Ay$. So, if x is tridiagonally majorized by y, then x is majorized by y. Intuitively, if x is tridiagonally majorized by y, then x may be obtained from y by a redistribution among consecutive components in y. (Remark: in contrast to majorization, tridiagonal majorization is not a transitive relation, an therefore not a preorder.)

It is natural to ask for a characterization of tridiagonal majorization in terms of linear inequalities involving the components of x and y. We now give such a result. In the theorem we consider a monotone vector $y \in \mathbb{R}^n$, so there are indices $1 \leqslant i_s \leqslant i'_s \leqslant n-1$ $(1 \leqslant s \leqslant p)$ with $i'_s \leqslant i_{s+1}-2$ and $y_i > y_{i+1}$ for $i_s \leqslant i \leqslant i'_s$ $(1 \leqslant s \leqslant p)$ and $y_i = y_{i+1}$ for all remaining indices $i \leqslant n-1$. We also define $i_{p+1} = n+1$ and the index set $I = \{1, \ldots, i_1 - 1\} \cup \bigcup_{s=1}^{p} \{i'_s + 2, \ldots, i_{s+1} - 1\}$.

Theorem 9. *Let $x, y \in \mathbb{R}^n$ be monotone, and let i_s, i'_s $(1 \leqslant s \leqslant p)$ and I be as above. Then x is tridiagonally majorized by y if and only if $x_i = y_i$ $(i \in I)$ and for $1 \leqslant s \leqslant p$*

(i) $\sum_{i=i_s}^{i'_s+1} x_i = \sum_{i=i_s}^{i'_s+1} y_i$,
(ii) $\sum_{i=i_s}^{k} x_i \leqslant \sum_{i=i_s}^{k} y_i \quad (i_s \leqslant k \leqslant i'_s)$,
(iii) $x_k \geqslant y_{k+1} + \frac{y_{k-1}-y_{k+1}}{y_{k-1}-y_k}\left(\sum_{i=1}^{k-1} y_i - \sum_{i=1}^{k-1} x_i\right) \quad (i_s \leqslant k \leqslant i'_s - 1)$.

If x is tridiagonally majorized by y and y is strictly decreasing, then there is a unique *tridiagonal doubly stochastic matrix A such that $x = Ay$.*

Proof. For given monotone x and y we consider the system $x = Ay$ where $A \in \Omega_n^{\mathrm{t}}$, i.e. (due to Proposition 1) $A = A_\mu$ with $\mu \in P_n$. In component form the system $x = A_\mu y$ becomes

$$x_i = \mu_{i-1} y_{i-1} + (1 - \mu_{i-1} - \mu_i) y_i + \mu_i y_{i+1} \quad (1 \leqslant i \leqslant n)$$

or equivalently

$$\mu_i (y_i - y_{i+1}) = \mu_{i-1}(y_{i-1} - y_i) + y_i - x_i \quad (1 \leqslant i \leqslant n), \tag{6}$$

where we define $y_0 = \mu_0 = y_{n+1} = \mu_n = 0$. This is a difference equation in the variables μ_i $(1 \leqslant i \leqslant n-1)$. Define $\alpha_i = y_i - y_{i+1}$ and $\Delta_i = y_i - x_i$ $(1 \leqslant i \leqslant n)$, so $\alpha_i \geqslant 0$. Then the system (6) decomposes into

$$\Delta_i = 0 \quad (1 \leqslant i \leqslant i_1 - 1)$$

and the following independent subsystems for $1 \leqslant s \leqslant p$

$$\begin{aligned} \alpha_{i_s}\mu_{i_s} &= \Delta_{i_s} \\ \alpha_{i_s+1}\mu_{i_s+1} &= \alpha_{i_s}\mu_{i_s} + \Delta_{i_s+1} \\ &\vdots \\ \alpha_{i'_s}\mu_{i'_s} &= \alpha_{i'_s-1}\mu_{i'_s-1} + \Delta_{i'_s} \\ 0 &= \alpha_{i'_s}\mu_{i'_s} + \Delta_{i'_s+1} \end{aligned} \tag{7}$$

and $\Delta_i = 0$ $(i'_s + 2 \leqslant i \leqslant i_{s+1} - 1)$. Here we have $\alpha_i > 0$ $(i_s \leqslant i \leqslant i'_s)$. Now, the subsystem (7) is consistent if and only if

$$\sum_{i=i_s}^{i'_s+1} \Delta_i = 0 \tag{8}$$

and then (7) has the unique solution μ_i $(i_s \leqslant i \leqslant i'_s)$ given by

$$\mu_i = \frac{\sum_{j=i_s}^{i} \Delta_j}{\alpha_i} \quad (i_s \leqslant i \leqslant i'_s).$$

In the solution set of (6) the remaining variables μ_i are free (i.e., when i is outside each set $\{i_s, \ldots, i'_s\}$). In summary, (6) is consistent if and only if $\Delta_i = y_i - x_i = 0$ $(i \in I)$ and (8) hold for $1 \leqslant s \leqslant p$. Moreover, the constraints $\mu_i \geqslant 0$ and $\mu_i + \mu_{i+1} \leqslant 1$ for each i (i.e., A_μ is doubly stochastic) translate into the remaining inequalities in the characterization of the theorem. Finally, if y is strictly decreasing, then $p = 1$ and each α_i is positive and therefore $\mu_1, \mu_2, \ldots, \mu_{n-1}$ are uniquely determined by (6). $\square$

We recognize conditions (i) and (ii) in the theorem as ordinary majorization conditions for certain subvectors of x and y. The proof of Theorem 9 also contains a complete description of the set of all tridiagonal doubly stochastic matrices A satisfying $x = Ay$. Finally, from the proof one also finds a characterization of tridiagonal majorization for possible nonmonotone vectors, but these inequalities are more complicated (as some α_i may be negative).

References

[1] A. Berman, N. Shaked-Monderer, Completely Positive Matrices, World Scientific Publ., 2003.
[2] R.A. Brualdi, Introductory Combinatorics, Prentice-Hall, 1999.
[3] R.A. Brualdi , H.J. Ryser, Combinatorial matrix theory, Encyclopedia of Mathematics, Cambridge University Press, 1991.
[4] R.A. Brualdi, An interesting face of the polytope of doubly stochastic matrices, Linear and Multilinear Algebra 17 (1985) 5–18.
[5] S. Karlin, H.M. Taylor, A First Course in Stochastic Processes, Academic Press, 1975.
[6] D. Kulkarni, D. Schmidt, S.-K. Tsui, Eigenvalues of tridiagonal pseudo-Toeplitz matrices, Linear Algebra Appl. 297 (1999) 63–80.
[7] S.-Z. Song, Y.-B. Jun, Minimum permanents of tridiagonal doubly stochastic matrices, Linear and Multilinear Algebra 50 (4) (2002) 301–306.
[8] M. Fiedler, An estimate of the nonstochastic eigenvalues of doubly stochastic matrices, Linear Algebra Appl. 214 (1995) 133–143.
[9] A.W. Marshall, I. Olkin, Inequalities: Theory of Majorization and Its Applications, Academic Press, 1979.
[10] C.D. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2000.